DOCUMENT RESUME

ED 235 845 JC 830 248

AUTHOR Olson, Margot A.; Martin, Diane
TITLE Assessment of Entering Student Writing Skill in the Community College.
PUB DATE Apr 80
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 7-11, 1980).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Community Colleges; *Essay Tests; *Holistic Evaluation; Interrater Reliability; Objective Tests; Remedial Instruction; Self Evaluation (Individuals); *Student Placement; Test Reliability; Two Year Colleges; Two Year College Students; *Writing Evaluation
IDENTIFIERS *Eastfield College TX

ABSTRACT

In fall 1978, a study was conducted at Eastfield College to determine the desirability of using a writing sample as part of the student assessment and advisement process. At an orientation session attended by 1,643 entering students, an entry-level assessment was conducted, consisting of a writing sample; an objective English composition test; the comprehension subtest of the Nelson-Denny Reading Test, Form D; and a self-assessment. Readers, who had been trained in the method and practice of holistic evaluation, provided recommendations based on the writing samples that students take credit English, developmental writing, or one of two combinations of credit and developmental courses. Based on data analyses performed on the scores of the entire student group and of the subgroup who enrolled in English and/or developmental courses, a study of interrater reliability, and interviews with readers and counselors, study findings included the following: (1) the writing sample was the most stringent placement instrument; (2) 39% of the students received the same recommendation from writing sample and objective test scores, and 35% from writing sample and self-assessment results; (3) 82% of the students followed recommendations to enroll in credit English, and 72% to enroll in developmental writing; (4) the objective test was the best single predictor of credit English grades; and (5) interrater reliability was reasonably high. (LAL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235845

ASSESSMENT OF ENTERING STUDENT WRITING SKILL IN THE COMMUNITY COLLEGE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. A. Olson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

X Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Margot A. Olson
Instructional Development Consultant
and
Diane Martin
Communications Instructor

Eastfield College
of the
Dallas County Community College District

The research reported herein was supported by a grant from Instructional Development Services at Eastfield College.

VC 830 248

Presented at the Annual Meeting of the American Educational Research Association, Boston, Massachusetts, April 7-11, 1980.

ASSESSMENT OF ENTERING STUDENT
WRITING SKILL IN THE COMMUNITY COLLEGE

Community college classes are filled by students representing all ages, backgrounds, and extents of academic ability. Such diversity is positive in that classroom populations closely resemble those of the community and exchange of ideas can be quite varied and stimulating. The diversity within student populations can become somewhat troublesome, however, when attempts are made to meet the individual needs of each student. It seems unreasonable to recruit non-traditional students, that is, those with little academic preparation or those returning to school after long absences, and then to treat them as if they were newly graduated high school students with an adequate academic preparation for college.

The philosophies upon which many community colleges develop their programs include the goal of providing positive educational experiences for all their students. Personal growth and development of the student is viewed as a goal which shares importance with that of pure academic achievement. To provide opportunities for personal growth rather than sending the student to certain failure, many students must be counseled into appropriate programs and some require non-credit work in reading, writing, and/or mathematics before beginning credit studies. To provide students with some knowledge on which to make decisions about readiness for credit college courses, many community colleges provide an entry-level assessment program for their students.

In a survey by Roueche and Snow (1977), it was reported that 83 percent of the community colleges provide some type of diagnostic assessment. In a later article, Roueche and Archer (1979) stated that "an increasing number of institutions are initiating assessment programs designed to place incoming students in the most appropriate 'track' and to provide developmental assistance to those

who need it (p. 15)." They go on to say that "there seems to be consensus among community college educators that entry-level assessment is necessary for successful guidance of students into appropriate programs (p. 16)."

The question facing community college educators appears not to be whether assessment is needed but rather what the entry-level assessment needs to contain and what type of information it needs to provide. Roueche and Snow (1977) reported usage of both published and locally designed instruments measuring both aptitude and personality factors. However, no suggested formula seems to exist for assessing probable success of entering community college students in credit offerings. It appears that each institution must determine its own procedure. The purpose of this report is to provide a description of the writing skills assessment which occurred at one community college.

Description of the College

Eastfield College is one of seven campuses of the Dallas County Community College District and has a head count enrollment of approximately 7500 students and a full-time equivalency enrollment of approximately 4000. Each fall, approximately 1600 students enter the college for the first time; each spring, new enrollment is around one-third the Fall figure. Within the District, new students are requested to attend an orientation where an assessment is given. There is little, if any, uniformity among the orientation and advisement procedures of the seven District campuses.

For several years, the entry-level assessment of writing skills at Eastfield College has included a short self-assessment questionnaire and an objective test in multiple-choice format. Time during the orientation session is limited, however, and the testing period has been made as short as possible. To perform the advisement function of orientation, immediate knowledge of test results has been necessary. The self-assessment and objective test scores have been used by counselors to advise students into credit English courses or developmental,

non-credit writing courses. Students themselves are permitted, however, to make the final decision about courses in which they enroll.

There has been some dissatisfaction among faculty across the college with the assessment procedures in that it appeared some students who needed developmental writing were not receiving it. This led the faculty to conclude that the objective test was not providing adequate information upon which students could make course decisions. From this, it was recommended that the objective test be replaced by or supplemented with a writing sample. Not only did faculty feel that a writing sample would provide a better indicator of writing ability than the objective test but they felt that students might be more willing to follow recommendations based on actual samples than on an objective test that seemingly had little relationship to writing skill.

Theoretical Framework

Use of a writing sample as part of entry-level testing provided two difficult logistical problems--administration and scoring. Earlier research on establishing reliability and validity of writing sample results (Coffman, 1966) indicated the need for two samples per student with each scored by five readers or three samples per student with each scored by three readers in order to be equivalent to a one-hour objective assessment of English composition. These writing and scoring recommendations were higher than could reasonably be established as part of the entry-level testing at Eastfield College. In fact, it seemed reasonable from administrative considerations to allow students to prepare only one 20-minute writing sample.

The later research of Breland and Gaynor (1979) suggested that experience in administering and evaluating writing samples has yielded improvements over the earlier results reported by Coffman. Breland and Gaynor felt that adequate reliability and validity could be established through careful collection and scoring of one sample per student and two readers per sample. Further,

Della-Piana, Odell, Cooper, and Endo (1976) concluded in their review of various ways to assess writing ability that "recent research and development in measurement demonstrates that we can achieve reliable and objective measures of nearly all important features of a piece of writing (p. 36)." In addition, they wrote that "the much-touted unreliability of human raters, their lack of agreement about grades or scores on papers, is a function of the specificity and clarity of scoring criteria and the amount of training (p. 36)." To provide an adequate indication of writing ability, however, Della-Piana, Odell, Cooper, and Endo recommended at least two writing samples from each student.

Considering the possible limitation of one 20-minute sample per student, the need for a very careful scoring process was obvious. The Coffman (1966) and Breland and Gaynor (1979) results were based upon "impressionistic" or holistic ratings of the writing samples. Della-Piana, Odell, Cooper and Endo (1976) stated that "anyone familiar with the composing process and the nature of written products would have to conclude that careful holistic judgments and descriptive procedures carried out on actual pieces of writing are the only way to obtain valid measures of writing (p. 36)." The holistic method of scoring yields one score reflecting the rater's general impression of the writing sample. This is contrasted to analytical scoring which includes multiple ratings on various specific qualities of the writing sample that are often combined in some weighted combination to yield a single score.

Holistic scoring was introduced as a highly reliable method for scoring essays by Godshalk, Swineford, and Coffman in a 1966 Educational Testing Service Monograph titled "The Measurement of Writing Ability." The College Entrance Examination Board has used the holistic method to score writing samples obtained from the English Composition Test with Essay. Diederich (1974) has described the process in detail as it might be used by classroom teachers. The holistic method is based upon the following four assumptions.

(1) The overall impact of a piece of writing is greater than the sum of its parts.

(2) Writing teachers are able to recognize good writing even though they often have difficulty describing, in concrete terms, what factors contribute to good writing.

(3) Raters may not agree on the relative importance of factors in analytic scoring yet will holistically rate papers in much the same way.

(4) The specific factors in analytic scoring cannot be removed from context and be independently judged; interaction among parts cannot be eliminated.

The holistic method assesses what is good about a piece of writing rather than what is wrong with it. Its success relies upon careful determination of standards associated with each score and thorough training of evaluators.

Since the holistic process had been shown to be highly reliable for use in evaluating writing samples and an overall impression of the writing sample was felt to be adequate for making placement recommendations, it was decided to use the holistic method at Eastfield College rather than the more diagnostic approach of analytical scoring. In addition, the holistic method had been found to be very rapid. A holistic rater can score from 30 to 60 writing samples per hour where the analytic approach requires considerably more time per sample. It was felt, too, that the writing faculty would have considerable difficulty determining the various factors and their weightings necessary for use of a more analytic procedure.

Importance of the Study

Proper selection of writing courses is a major problem facing students entering community colleges. Procedures cannot be so rigorous that they cause students to forego college in this period of declining enrollments nor so lax that they provide students no information for making enrollment decisions. Therefore, community college personnel must begin sharing their successes and failures in the area of entry-level assessment.

Some factors associated with the importance of such studies as the one reported here are the following:

(1) greater reliability and validity of writing assessment procedures for entering community college students,

(2) better placement of students in courses from which they will receive maximum benefits,

(3) reduction of college-wide complaints that students are unable to adequately write term papers and essay exams,

(4) empirical determination of cut-off scores for recommending various courses to students, and

(5) development of a model for implementing the holistic reading approach for scoring writing samples of entering students.

It was for these reasons that financial support to conduct a study of various writing assessment instruments was granted by the Eastfield College administration through the Instructional Development Services grant fund.

METHOD

Student Population

Students who planned to enter college for the first time during the Fall, 1978 semester at Eastfield College were asked to attend an orientation and advisement session during the summer preceding enrollment. The sessions were set up throughout the summer during day and evening hours to allow students flexibility in coming at convenient times. During the sessions, students learned about the college, completed entry-level assessment instruments, and determined a tentative listing of courses for enrollment in the fall. The entry-level assessment included a writing sample based upon one of seven topics; an objective test of English composition; the comprehension subtest of the Nelson-Denny Reading Test, Form D; and a self-assessment titled Encounter with Eastfield. The orientation and advisement session for the Fall, 1978 semester was attended by 1643 students. Because of the experimental status of the writing sample assessment, students did not receive knowledge of their writing sample scores. According to the usual process, they did receive knowledge of their scores on the objective test, Nelson-Denny Reading Test, and self-assessment.

Data Sources

Writing sample. To eliminate the possibility of a student preparing a writing sample in advance of the testing session and for purposes of try-out, seven topics were selected for administration during the Summer of 1978. Topics were selected so that nearly any individual would have had some experiences relevant for preparing a short writing sample. A one-sentence summary of each topic is given below.

(1) Turn back time and describe your life as you might have lived at some earlier period in history.

(2) Describe how your life is similar to or different than your parents' lives.

(3) Discuss the reaons why you feel the divorce rate is on the increase.

(4) Describe the kind of person you would like to be and explain why.

(5) Discuss the factors you feel make television a positive aspect of life.*

(6) Describe the good you see in materialism, that is, people's desires to buy and own better things.*

(7) Describe the positive aspects of sports in our society.*

Each topic was preceded by the following instruction.

> During the next twenty minutes, plan and write an essay on the following topic. You will not be given credit for an essay on any other topic.

The topic description was followed by the instruction given below.

> In your writing, be specific. Carefully express your thoughts clearly and logically with examples, illustrations, and experiences.

Testing personnel were requested to systematically vary the topics for the writing sample and the order of administering the writing sample and objective test. Testing center personnel were requested to administer topics in rotating fashion to each successive group of students; they were requested to reverse the order of the writing sample and objective test for every other group of students. Students were requested to identify themselves by placing only their social security numbers on their writing samples.

Writing samples were scored on a scale of 1 (poor) to 4 (good). The scores were related to four placement opportunities available to students. The following criteria were established to provide all readers with the same general grading standards.

4. Credit English Recommendation. Organized, varied sentence structure, creative, in-depth reaction to topic, good mechanics, and no glaring writing errors.

---

*Used by permission of Educational Testing Service.

3. Credit English and One Credit Hour Developmental Writing Recommendation. Sticks to topic, not quite as detailed and organized as a writing sample given a "4" rating, some mechanical errors, and could survive in credit English.

2. Special Combination Course of Credit English and Three Hour Developmental Writing Recommendation. May or may not stay on topic, no depth or insight into topic, simplistic but understands topic, and shows need for vocabulary and mechanical improvement.

1. Three Credit Hour Developmental Writing Recommendation. Definite problems with topic, language, and mechanics; simplistic approach to topic.

Readers were advised that writing samples which deviated significantly from the topic should receive no more than a score of "2."

Objective test. The objective test of English composition was a locally produced instrument composed of 40 multiple-choice items. The items were written in two formats, each composing half the test. Two example items follow.

Tomorrow, we walked (A) to the park at the corner of (B) First Avenue and (C) Main Street. No Error (D)

She and her went on a vacation to Colorado.

A. She and I
B. They
C. We
D. No Change

Time limit for the test was 20 minutes. Items tested such writing skills as subject and verb agreement; using complete sentences; placing modifiers; appropriate subordination and coordination of phrases; and using appropriate pronouns, modifers, and verbs. The items did not test spelling, capitalization, or grammatical terms. For placement purposes, the following score ranges were used to recommend the four writing course possibilities.

31-40 = 4. Credit English
23-30 = 3. Credit English/One Hour Developmental Writing
16-22 = 2. Credit English/Three Hour Developmental Writing
0-15 = 1. Three Hour Developmental Writing

Self-assessment. The self-assessment form included the following three items relevant to writing skill.

I have completed _ senior English _ honors English
_ college English.

My average grade in English has been _A _B _C _D _F.

My abilities and skills in writing are _ good _ average
_ poor.

For use in this study, the third item only was included as an indicator of self-assessment. This yielded scores of 1 (poor) to three (good). For comparisons with the writing sample and objective test, the following placement recommendations were used with responses to the third self-assessment item.

3 = 4. Credit English
2 = 3. Credit English/One Hour Developmental Writing
1 = 2. and 1. Developmental Writing with or without Credit English

Other variables. Other variables, available for comparison and possibly relating to success in college writing courses selected by entering community college students, were the comprehension score of the Nelson-Denny Reading Test, first semester writing course selection and grade, age, and sex.

Nelson-Denny Reading Test, Form D. The comprehension section of this test consists of 36 items which measure ability to read short passages and answer questions about the content. The raw score was used for data analysis.

Writing Course Selection and Grade. For those students who enrolled in credit English and non-credit developmental writing during the Fall, 1978 semester, course selection and end of semester grade were obtained.

Age. The age in years at the time of entry to the college was obtained.

Sex. The sex of entering students was obtained and coded "-1" for males and "1" for females.

Procedure

During the first week of the Fall, 1978 semester, all professional and para-professional employees of the college were invited to attend a half-day workshop for training in the method of holistic reading. The participants included primarily faculty from English and developmental teaching areas but included, also, participants from other teaching disciplines, counseling, and

learning resources. From a group of approximately 165 possible participants, approximately 40 persons elected to attend the workshop. Training included a review of the method and actual practice in using the method.

After the training session, all interested persons were invited to become members of the pool of readers. The resultant pool of readers included 12 English instructors, 3 developmental studies instructors, 2 social science instructors, 3 para-professional counselors, and 3 learning resource personnel. Schedules were determined to bring groups of readers together for scoring the writing samples collected during the past summer's orientation sessions. The reading groups consisted of three persons; all groups included at least one English or developmental writing instructor. A process for scoring and recording scores was devised from the Educational Testing Services methodology to fit local conditions.

Writing samples were prepared for scoring by arranging them in groups of approximately 60 examples of student response to the same topic and by giving each a unique identification number. To avoid marking scores on the writing samples, scoring sheets were prepared for each reader. To provide a starting point for reader discussion of standards as they related to particular topics, writing samples thought to be typical of each score, called range-finders, were identified.

Groups of readers were seated on three sides of a table and were given the following items: three score sheets, three pencils, approximately 60 writing samples of the same topic, four range-finders, and a copy of the score criteria and grading standards. The steps below describe the process followed to determine a score for each writing sample.

(1) Range-finders and grading standards were discussed to achieve group concensus on the meaning of each score for the particular topic.

(2) Six writing samples were distributed around the table and rotated among the readers until all readers determined an independent score for each sample. As readers determined scores for writing samples, they recorded each on their score sheets.

(3) Readers compared scores given to each writing sample to determine a single, final score.

(4) When scores differed among readers, discussion was undertaken to achieve concensus. Generally, the majority score was given when differences occurred between scores of "1" or "2" and scores of "3" or "4." When the split involved non-adjacent scores or a "2" and "3" split (point at which three-hour developmental course was recommended), discussion was more thorough. When a decision was reached, the score was recorded on the score sheets.

(5) If readers were unable to agree on a single score, the writing sample was set aside to be given later to another group of readers. A "?" was entered on the score sheets.

(6) The process was repeated with groups of six writing samples until the entire set was scored or one hour passed.

It was found that groups scored an average of 58 writing samples per hour.

To provide an indication of inter-rater reliability, six writing samples were randomly selected from each of the 30 groups of scored papers. These were rescored, following an identical procedure, by nine of the original group of readers reassigned to three groups of three readers. In addition, all readers and counselors were interviewed to gather information about attitudes toward and perceptions of the process. An interview schedule was prepared for each group and administered by the two authors of this report.

Data Analysis

In order to make a recommendation concerning the desirability of using the writing sample as part of the advisement process, data analyses were performed as they relate to the entire group of 1643 students, the subgroup of students who enrolled in English and/or developmental writing courses, the study of inter-rater reliability, and the survey of reader and counselor attitudes toward the process. Analyses using writing sample scores were performed both across topics

and by topic where appropriate. Most descriptive results were tabulated using the program CROSSTAB (Olson, 1978), and the regression analyses were computed using the program CORRAL (Olson, 1979).

RESULTS

Entire Population

The number of students given each placement recommendation for each writing assessment is given in Table 1. The self-assessment instrument resulted in the least number (11%) of recommendations of developmental writing (i.e., recommendations 1 & 2). The objective test was next with 35% recommended to developmental writing. To a similar extent, writing sample scores indicated need for developmental writing by 54% of the students. The writing sample resulted in the least recommendations (5%) to English 101 without a developmental course combination. The objective test and self-assessment resulted in 25% and 31%, respectively, receiving a recommendation of "4." It seems, therefore, that the writing sample was the most stringent of the placement instruments.

TABLE 1

Numbers of Students Receiving
Each Placement Recommendation

Writing Assessment		Placement Decision 1	2	3	4	Total
Writing Sample	F	175	713	678	77	1643
	%	10.65	43.40	41.27	4.69	100.00
Objective Test	F	165	352	591	374	1482
	%	11.13	23.75	39.88	25.24	100.00
Self-Assessment	F	121		638	334	1093
	%	11.07		58.37	30.56	100.00

To determine the similarity of recommendations resulting from the writing sample and from the objective test and self-assessment, comparisons are given in Table 2. From Table 2, it can be observed that nearly the entire range of writing sample scores was received by students representing all categories of objective test score and self-assessment. Only 39% of the students received the same recommendation from writing sample and objective test scores and 35% of the students received the same recommendation from the writing sample and self-assessment.

TABLE 2

Comparison of Placement Recommendations
from the Three Writing Assessments

Writing Sample Score		Objective Test Recommendation 1	2	3	4	Self-Assessment Recommendation 1 & 2	3	4
1	F	60	69	19	3	33	62	16
	%	4.05	4.65	1.28	0.20	3.02	5.67	1.46
2	F	91	194	276	81	59	295	109
	%	6.14	13.09	18.62	5.47	5.40	26.99	9.97
3	F	14	86	276	247	26	264	171
	%	0.94	5.80	18.62	16.67	2.38	23.38	15.64
4	F	0	3	20	43	3	17	38
	%	0.00	0.20	1.35	2.90	0.27	1.56	3.48
Total	F	165	352	591	374	121	638	334
	%	11.13	23.75	39.88	25.24	11.07	58.37	30.56

Means, standard deviations, intercorrelations, and numbers of respondents are given in Table 3 for the four assessment instruments. All correlations were statistically significant ($p < .01$). As might have been expected, the largest correlation occurred between the scores from the two multiple-choice instruments, the objective test and the Nelson-Denny Test.

TABLE 3

Descriptive Statistics for the Four
Assessment Instruments

Assessment Instrument	N	Mean	S.D.	Correlations and Numbers of Respondents 1	2	3	4
1. Writing Sample	1643	2.40	0.76		1482	1093	1446
2. Objective Test	1482	24.97	7.28	0.53		1058	1417
3. Self-Assessment	1093	2.20	0.61	0.28	0.33		1032
4. Nelson-Denny	1446	17.65	6.14	0.41	0.69	0.28	

For each topic of the writing sample assessment, the means, standard deviations, numbers of respondents, and correlations with the objective test, self-assessment, and Nelson-Denny are given in Table 4. To provide an indication of

TABLE 4

Descriptive Statistics for Each Topic of the Writing Sample Assessment

Writing Sample Topic	N	Mean	S.D.	Objective Test		Self-Assessment		Nelson-Denny	
				N	Correlation	N	Correlation	N	Correlation
A	200	2.56	0.91	178	0.54	136	0.30	176	0.47
B	231	2.52	0.75	204	0.50	130	0.28	210	0.37
C	306	2.37	0.75	278	0.53	204	0.35	269	0.51
D	279	2.43	0.71	253	0.48	175	0.17	233	0.39
E	121	2.40	0.71	113	0.56	63	0.38	112	0.39
F	298	2.26	0.72	265	0.53	247	0.18	254	0.41
G	208	2.33	0.72	191	0.62	138	0.44	192	0.41

the effect of order of administration of the writing sample and objective test, means, standard deviations, numbers of respondents, and correlations of writing sample with the other three instruments are given in Table 5. All correlations in Tables 4 and 5 are statistically significant ($p < .05$). In reviewing the means and standard deviations in Tables 4 and 5, few differences between topics or order of administration are apparent. With the exception of self-assessment values, perhaps, the correlations, too, are quite similar. It seems that Topics A and B are slightly easier than the other topics, that the range of scores for Topic A is somewhat larger than the others, and that self-assessment has less relationship with writing sample scores resulting from Topics D and F.

TABLE 5

Descriptive Statistics by Order of Administration for the Writing Sample Assessment

Order of Administration	N*	Mean	S.D.	Objective Test		Self-Assessment		Nelson-Denny	
				N	Correlation	N	Correlation	N	Correlation
1	380	2.40	0.73	340	0.58	259	0.33	335	0.44
2	1181	2.40	0.75	1074	0.52	769	0.27	1040	0.41

*Order of Administration was not coded for 82 students.

To determine if the differences observed in Tables 4 and 5 are statistically significant, results of a 7 x 2 Analysis of Variance are presented in Table 6. None of the F-statistics were statistically significant ($p < .05$). Topic and order of administration, therefore, did not appear to affect the scores on the writing samples. There is no evidence to imply that topics used for this study or order of administration affected the writing sample scores. Because of lack of evidence of differences, the seven topics would seem to be equivalent and the order of administration of the two instruments would seem not to matter.

TABLE 6

ANOVA of Writing Sample Topic and Order of Administration

Source	df	SS	MS	F
Topic	6	4.797	0.800	1.454
Order	1	0.425	0.425	0.774
Topic X Order	6	2.684	0.447	0.813
Error	1547	850.635	0.549	

Writing Course Students

Information about enrollment in writing courses is presented in Table 7. The table includes the number of students who enrolled in each placement option. In addition, the percentage of students whose placement recommendation matched their enrollment choice is given for the writing sample, objective test, and self-assessment (due to missing data for some enrollees, these values are given as percentages). Because of the small number who actually enrolled in options 2 and 3, options 1 and 2 and options 3 and 4 have been combined to represent enrolling in the three-hour developmental course versus not enrolling in the three-hour developmental course. The percentages for the two combined options represent students receiving scores of either "3" or "4" and either "1" or "2."

TABLE 7

Enrollment in Writing Courses

Placement Options	Number Enrolled	Percent Who Enrolled According to Recommendation		
		Writing Sample	Objective Test	Self Assessment
4. Credit English	557	5.30%	36.35%	44.78%
3. Credit English/1 hr. DW	11	27.27%	63.64%	55.55%
Total 3 & 4	568	55.84%	81.73%	97.25%
2. Credit English/3 hr. DW	19	73.68%	36.84%	--
1. Three-Hours DW	161	21.79%	26.62%	--
Total 1 & 2	180	82.23%	71.67%	27.19%

For deciding whether or not to enroll in developmental writing, the objective test seems to be best at representing the enrollment decisions made. Of course, this recommendation was available to students prior to their actual enrollment and, therefore, reflects enrollment of those students who followed the recommendation. From the percentages of 82% in credit English and 72% in developmental writing, it is apparent that many students do heed the placement recommendation. Also, it seems that many students who enroll in developmental writing have over assessed their writing skill on the self-assessment instrument. Since only 27% of the developmental writing students rated themselves as "poor," the remaining 73% apparently rated their skills as "average" or "good." It appears that many students who over assessed themselves on the self-assessment decided to heed the recommendation based on the objective test. The writing sample scores were unavailable at the time enrollment decisions were made. The results based on writing sample scores as compared to objective test and self-assessment would indicate, however, less consistency between enrollment choice and recommendation for students enrolling in option 3 or 4. As concluded earlier, the writing sample appears to set a higher standard for placement recommendations.

Regression analyses were performed to determine which test scores were the best predictors of performance in a writing course. Because of the small numbers of students enrolled in the credit/non-credit combinations, predictive analyses were done only with results of credit English and developmental writing students as defined by recommendations 1 plus 2 and recommendation 4. Since credit English followed developmental work in recommendation 2, those receiving recommendation 2 were included only with developmental students. Because of the interaction of the one-hour developmental course running simultaneously with the English course, students in option 3 were eliminated from the regression analyses. Students who withdrew (W) or received grades of incomplete (I) or progress (P) were not included in the regression analyses. Dependent variable for the regression analyses was, therefore, a numerically coded performance grade of A, B, C, D, or F.

Grade distributions for students enrolled in credit English and students enrolled in developmental writing are given in Table 8. A very high percentage of "A's" was given in the developmental writing course. This was no doubt due

TABLE 8

Grade Distributions for Credit English and Developmental Writing Students

COURSE		GRADE								TOTAL
		A	B	C	D	F	W	P	I	
English										
Total Group	F	113	144	107	25	29	122	2	15	557
	%	20.29	25.85	19.21	4.49	5.21	21.90	0.36	2.69	100
Performance Grades Only	F	113	144	107	25	29	--	--	--	418
	%	27.03	34.45	25.60	5.98	6.94	--	--	--	100
Developmental Writing										
Total Group	F	79	31	4	1	2	40	19	4	180
	%	43.89	17.22	2.22	0.56	1.11	22.22	10.56	2.22	100
Performance Grades Only	F	79	31	4	1	2	--	--	--	117
	%	67.52	26.50	3.42	0.85	1.71	--	--	--	100

to the mastery approach used to develop specific writing skills. The high concentration of "A's" reduced the variance of the criterion variable for the regression analyses of the developmental writing students. The grades of the English students were spread more evenly across the "A, B, C" range and resulted in more variance for the criterion variable for that group.

Descriptive statistics for both credit English and developmental writing students are presented in Tables 9 and 10 respectively. In addition to the four test scores are values for age and sex, the two other variables available for regression analyses.

TABLE 9

Descriptive Statistics for Credit
English Students Earning a Performance Grade

Variable	N	Mean	S.D.	Correlation and Number of Respondents						
				1	2	3	4	5	6	7
1. Writing Sample	418	2.61	0.72		413	273	397	418	418	418
2. Objective Test	413	28.38	5.63	0.39*		273	397	413	413	413
3. Self-Assessment	273	2.41	0.58	0.33*	0.36*		273	273	273	273
4. Nelson-Denny	397	20.06	5.54	0.29*	0.56*	0.23*		397	397	397
5. Age	418	20.63	5.89	0.05	0.18*	0.00	0.01		418	418
6. Sex	418	-0.01	1.00	0.17*	0.19*	0.08	0.03	0.09		418
7. Grade	418	2.69	1.14	0.24*	0.45*	0.21*	0.30*	0.21*	0.17*	

* $p < .01$

TABLE 10

Descriptive Statistics for Developmental Writing Students
Earning a Performance Grade

Variable	N	Mean	S.D.	Correlation and Number of Respondents						
				1	2	3	4	5	6	7
1. Writing Sample	117	2.01	0.61		116	74	112	117	117	117
2. Objective Text	116	19.14	4.85	0.34*		74	112	116	116	116
3. Self-Assessment	74	1.84	0.62	0.23†	0.05		74	74	74	74
4. Nelson-Denny	112	14.17	4.36	0.28*	0.48*	0.14		112	112	112
5. Age	117	21.70	5.17	0.33*	-0.18†	-0.34*	-0.08		117	117
6. Sex	117	-0.28	0.96	0.17	0.05	0.08	0.06	0.14		117
7. Grade	117	3.58	0.75	0.22†	0.13	-0.16	0.02	0.27*	0.16	

* $p < .01$
† $p < .05$

Looking at correlations of the six predictor variables with grade, it is apparent that relationships between scores differ between the two groups of students. The objective test with a correlation of 0.45 is apparently the best single predictor of grade for English students. This is followed by the Nelson-Denny with a correlation of 0.30 with grade. The writing sample correlation with grade is 0.24, only slightly higher than self-assessment and age correlations. For developmental writing students, the best predictor of grade is age (0.27) followed by the writing sample (0.22). The objective test correlation with grade is 0.13 and the Nelson-Denny is 0.02. It may be that reading ability affects English performance because of combined reading and writing assignments but has little relationship with the more skill oriented developmental course.

To determine if combinations of variables would increase the relationship with grade, the results of multiple regression analyses are presented in Table 11. The combined prediction of the writing sample and objective test, the combined prediction of the two variables with the highest correlation with grade, and the combined prediction of all six variables is given.

TABLE 11

Prediction of Course Grades
in English and Developmental Writing

	Predictor Sets	R's and Standard Regression Weights			
		English		Developmental Writing	
1.	Writing Sample	R=.464	(.241)	R=.235	(.216)
	Objective Test		(.352)		(.060)
2A.	Objective Test	R=.460	(.280)	--	--
	Nelson-Denny		(.295)	--	--
2B.	Writing Sample	--	--	R=.333	(.216)
	Age	--	--		(.238)
3.	Writing Sample	R=.523	(.212)	R=.395	(.189)
	Objective Test		(.173)		(.077)
	Self-Assessment		(-.128)		(.219)
	Nelson-Denny		(.199)		(-.011)
	Age		(.152)		(.157)
	Sex		(.168)		(.164)

As far as placement recommendations are concerned, the prediction of English grade is probably more relevant for consideration than is prediction of developmental writing grade. In making a course recommendation to a student, it is with the intent of maximizing the probability of success in the credit offering. For this reason, the best predictors for English should be given primary consideration in selecting the appropriate entry-level testing format.

From Table 11, it can be observed that the combinations of objective test with writing sample (R = 0.46) and Nelson-Denny (R = 0.46) do little to improve the correlation of objective test alone with grade (r = 0.45). In fact, the addition of all the predictor variables to the model (R = 0.52) does little to improve prediction over the objective test alone. Based on these results, the use of the objective test alone is probably the most efficient predictor of success in credit English.

For developmental writing, prediction using the combination of writing sample and age (R = 0.33) is somewhat better than age alone (r = 0.27). The full set of predictors (R = .40) accounts for a similar increase beyond the best combination of two predictors. Based on these results, it appears that older students perform better in developmental writing than younger students and that those older students who begin with more writing skill as measured by the writing sample can be expected to do well in developmental writing.

Reliability Analysis

Writing sample. To provide an indication of the consistency of scores across groups of readers, a comparison of the two scores for the 180 writing samples which were rescored is given in Table 12. A total of 73.89% of the scores given by the two groups were the same. Since 8.89% of the scores were higher and 17.22% of the scores were lower from the second reading, the second group of readers tended overall to be more critical of writing ability than was the first group. It is important to note that there were no deviations of more

than one point and that nearly 18% would have received different recommendations concerning need for developmental writing.

TABLE 12

Comparison of the Two Scores for Writing Samples Included in the Reliability Study

First Scoring		Second Scoring 1	2	3	4	TOTAL
1	F	12	4	0	0	16
	%	6.66	2.22	0	0	8.89
2	F	5	64	12	0	81
	%	2.78	35.56	6.66	0	45.00
3	F	0	20	53	0	73
	%	0	11.11	29.44	0	40.56
4	F	0	0	6	4	10
	%	0	0	3.33	2.22	5.56
TOTAL	F	17	88	71	4	180
	%	9.44	48.89	39.44	2.22	100

The means and standard deviations from both scorings and the correlation between the two scores are presented in Table 13 for each of the topics.

TABLE 13

Descriptive Statistics by Topic for the Two Scores of the Writing Samples Included in the Reliability Study

Topic	N	Mean 1	Mean 2	SD1	SD2	Correlation
A	24	2.583	2.458	.862	.763	.796
B	30	2.633	2.533	.547	.499	.473
C	30	2.379	2.448	.665	.497	.690
D	30	2.690	2.621	.835	.925	.767
E	12	2.167	2.000	.553	.408	.750
F	30	2.133	1.800	.743	.644	.678
G	24	2.291	2.291	.889	.735	.827
TOTAL	180	2.426	2.343	.733	.679	.733

All correlations were significant ($p < .01$). The means, standard deviations, and correlations from the two scorings were similar for the most part. The means for Topic F revealed the largest difference of 0.333 yet the correlation

was still relatively high (r = 0.68). The correlation for Topic B (0.47) however, was considerably outside the range of the other six correlations (0.68-0.83). Although more information would be useful for making a judgment, it might be the Topic 2 yields writing samples where agreement among readers is harder to achieve. In fact, 8 of the 32 split decisions between scores of 2 and 3 occurred for Topic B. If the split decisions had occurred proportionately across topics, Topic B should have 5.33 such splits. For Topic B, more discussion in setting standards is probably needed to improve consistency of ratings across groups.

Overall, the consistency between groups of readers seems reasonably high. Although the correlation of 0.733 across all topics could be improved, it is important to note that this was a new process being tried by this group. Improvement over time could be expected with the result of higher agreement overall and fewer inconsistencies between the critical scores of 2 and 3.

Objective test. A Kuder-Richardson Formula 20 (KR-20) estimate of reliability was computed for the objective test of English composition. Descriptive statistics and reliability coefficient for a randomly selected subset of 279 of the 1482 students whose scores on the test were available is given in Table 14.

TABLE 14

Descriptive Statistics and Reliability Estimate for the Objective Test

Statistic	Value
Mean	25.3
Median	26.0
Mode	24
Range	36 (4 to 40)
Standard Deviation	7.1
KR-20 Reliability	0.857

The reliability of 0.857 is adequate for a locally produced test. This estimate is similar to a KR-20 reliability coefficient of 0.863 computed on a sample of

150 drawn from entering students of the previous semester. The similarity of the coefficients from two administrations adds to the claim of adequate test reliability for the objective test.

Interviews with Readers and Counselors

Readers. The questions asked the readers were grouped in nine distinct areas containing from one to seven items. Questions were asked in the following areas: workshop, holistic reading process, scoring system, scoring sheets, topics for writing, time element, benefit of process, standards, and readers. A brief summary of the responses of the 23 readers is given below for each of the nine areas of questioning.

In general, readers felt the workshop was effective in training them in the process of holistic evaluation of writing samples. Comments revealed a dislike for the detail and length of time used to present the background information about the method. There was general agreement that actual practice in using the technique was very valuable. Nearly all the readers recommended a workshop for groups planning to implement holistic reading.

There was unanimous approval of the process developed for evaluating the writing samples. In general, it was felt that groups of three readers could work independently without a fourth person coordinating the process. There was the feeling, however, that someone needed to have responsibility for coordinating the different groups, handling problems, and keeping records.

The scoring system of "1," "2," "3," and "4" was clear to all but one reader. The person who found it unclear felt that some samples did not fit distinctly into a category. Asked whether a six point scale would have been more useful, 16 respondents answered "no." All respondents felt the process for resolving disagreements was acceptable. In scoring writing samples that deviated from the topic, 14 respondents felt that an automatic recommendation to developmental writing was unreasonable. It was felt that writing skill should determine the

score except, perhaps, wide topical deviation should not receive a score of "4."

There were no suggestions for improving or difficulties with the sheets used to record the scores. When asked if they liked the flexibility of discussing essays where disagreement occurred, all answered "yes." Response was quite varied as to who should determine the score on writing samples where the original group could not agree. Ten of the respondents felt that one individual outside the group could determine the score; eight felt that another group should determine it.

When asked about topic choices and possible deletions of topics, there was little agreement. It might be that a small group rather than all readers should suggest topics for pilot administrations. Most felt that seven topics was an appropriate number and that two or three topics would be too few. When asked if readers should write essays on the topics, 17 responded "no"; 3, "yes"; and 3, "doesn't matter."

Most readers felt that one hour was an appropriate block of time for scoring writing samples. There was general agreement that the room used for scoring was uncomfortable and provided distractions. The scheduling process for setting up times and groups of readers was reported to be generally acceptable.

Of the 23 readers, 20 felt the writing sample to be an effective screening technique and 3 had no opinion. All but one, who did not care, indicated that the process should be continued. There were many additional benefits of the process listed. These included interaction with other faculty, familiarity with student writing ability, recognition of various student attitudes expressed in the writing samples and professional development.

The standard that created the major concern among readers was the one relating to scores for essays that deviated significantly from the topic. It was felt by some that an automatic "2" was too harsh a penalty for some students who deviated

from the topic. A few others felt that the standards for scores of "3" and "4" were too high.

The qualifications suggested for readers was quite varied. There were no specific complaints about any of the persons who participated in the pilot project. Of the 23 readers, 14 felt there was a noticeable difference among groups. The difference was primarily in regard to speed of scoring. Nearly all respondents felt that it was appropriate to include readers representing areas other than writing, that this permitted exposure to varied philosophies, and that it provided additional support for the process.

In summary, the readers were, in general, pleased with the way the holistic reading project was implemented. The responses indicated a need for review of the topics and standards to make these clearer and more consistent. The actual mechanics of the process could apparently be continued as originally designed.

Counselors. The counseling staff was involved with the holistic reading process as a part of their advisement duties. Once scores were determined, it was planned that counselors would aid students in interpreting their scores and in understanding the options available to them. Because of this responsibility, opinions of the counselors concerning the writing assessment process were sought. At the time the data for this study were collected, there were 11 counselors and a director of counseling on the staff. To gather opinions about the holistic reading process, six members of the counseling staff, who had had some involvement with the project, were interviewed. A list of questions, similar to those asked the readers, was developed to gather information about the counselors' impression of the holistic reading process as an entering student assessment tool. The questions were grouped in five distinct areas containing from one to four items. A brief summary of responses is given below for each of the five areas: workshop, topics, standards, readers, and impressions of the process.

The workshop on holistic scoring was attended by five of the six counselors who were interviewed. They reacted favorably to the concept of the writing sample as an assessment tool but felt, like the readers, that the workshop format could have been better.

In reviewing the topics, the counselors were favorable for the most part. Specific comments were made that the divorce topic might cause overreaction by those students recently experiencing divorce and that topics avoiding possible "canned" responses would be best.

The standards were viewed as generally acceptable by the counselors. There was some concern that some terms used to define the standards should be expanded upon. Several counselors felt that the standards were too high for scores of 3 and 4.

Responses about reader qualifications varied from having two years of English to having at least a master's degree. There was uniform response that readers should represent different areas of the campus rather than just writing faculty. It was felt that writing is a campus concern.

Counselors indicated that the process did provide some benefits to the student and revealed much to the readers about the skill and attitudes of the student population. Three counselors believed that the writing sample was a better assessment technique than the objective test. There was a general reaction that students needed to receive immediate feedback if the writing sample is to be effectively used in advising new students. There was a general feeling that results of the pilot study would be needed before a final conclusion of the benefits of the process could be made.

DISCUSSION

The purpose of the study reported here was to gather information to be used in making a decision about the best entry-level assessment for Eastfield College students. Specifically it was to look at the current process using an objective test and self-assessment and to consider using a writing sample as the basis for recommending credit English and non-credit developmental writing courses to students. It seems critical in this period of recruiting students who may not be adequately prepared for college credit courses to provide the best available information to them for making appropriate course selections. For a first semester student, rewarding academic experiences can be crucial in completing the semester and returning to the college for future courses.

The statistical results of this paper form only one component of the basis for designing the entry-level assessment program. As Breland and Gaynor stated in their study of direct and indirect measures of writing ability, "indirect measures lack face validity and credibility among members of the English profession and educators generally, and they tend to deliver a message to students that writing is not important (p. 127)." It may be that a writing sample, no matter what its relationship to other variables, is needed as a part of entry-level assessment, not so much for its use as a predictor but because it points out the value of writing skill in attempting college credit courses.

Considering the comparisons of topics and order of administration, reliability studies, and attitudes of readers and counselors, no major problems seemed to emerge that would hinder the use of the writing sample as part of the assessment of entering student writing skill. Topics seemed quite similar in means, standard deviations, and correlations with other tests. The administration of the objective test prior to the writing sample, a possible "warm-up" exercise, seemed to have no effect. Consistency of scores for the same writing

sample between two independent but similarly trained groups of readers was reasonably good. With further refinement of topics and greater clarification of standards, consistency of rating across groups of readers would probably improve. The reliability of the objective test was adequate and like the writing sample, could possibly be improved slightly with detailed analysis of test items. Attitudes of readers toward the process was excellent. The scoring format, except for room accomodations, received no criticisms. Counselors appeared very open to using a writing sample as an advisement tool.

From the results of comparisons of the writing sample, objective test of English composition, and self-assessment, it is obvious that the three approaches are measuring different aspects of writing ability. Some students, in fact, received the highest recommendation from one measure and the lowest recommendation from another measure. Depending on the instrument, the percentages of students receiving each recommendation varied considerably, also, with differences between instruments as large as 30%. For some students, the differences in recommendations from the various instruments were minimal; for others, the variety of scores would have made a great difference in course selection as well as self-perception.

The recommendation of whether or not to take developmental writing is the critical decision to be made by the students. Results showed very low enrollment in the two credit English and developmental course combinations. Only 4% of the 748 students who enrolled in a writing course selected recommendation 2 or 3 on the objective test. Had they followed the objective test recommendation, 64% should have enrolled in a combination offering. The split, however, for those receiving a "2" and not enrolling in a combination course was approximately 50% to credit English only and 50% to developmental writing only. This almost implies that students receiving "2's" tossed a coin to make their choices. Of those receiving a "3" and not enrolling in a combination, 86% enrolled in credit English

only. It might be wise to recommend either developmental writing or no developmental writing as a first suggestion followed by the recommendation of the combination, if appropriate, as a consideration. At least, then, those students receiving "2's" would hear the need for developmental writing first and, if they could not or would not enroll in the combination, hopefully would enroll in developmental writing rather than English.

The choice of a criterion is difficult in a prediction study. It is almost because of a lack of anything better that grade in writing course was used in this prediction study. In considering the design of assessment procedures for entering students, it seemed reasonable to consider credit English and developmental students separately. It is, after all, success in credit courses that is critical to course selection advisement. Fortunately, some students representing each score on the three instruments chose to enroll in credit English. Considering the correlations and regression analyses, therefore, based on credit English enrollees, the best predictor of English grade was the objective test. Increases in prediction from adding other variables to the prediction equation were so minimal that, in practicality, it probably would not be worth the effort to combine and explain the more complex scores. The writing sample and even the Nelson-Denny Reading Test scores are probably best used to individually counsel those students who feel unsure about selections made on the basis of the objective test.

In considering any conclusions drawn from the results presented here, it should be remembered that the study was conducted using only one group of students, those who enrolled in the Fall of 1978. To give greater credibility to any conclusions drawn, it is necessary to study more groups at different times. These results, however, provide some information that can be built upon over time and with similar experiences from other community college groups. In summary, the following conclusions can be drawn from the study.

1. Students representing a wide range of writing ability enter Eastfield College.

2. The three writing assessment instruments, the writing sample, the objective test, and the self-assessment, result in quite varied course recommendations for many students.

3. The writing sample was the most stringent of the placement instruments.

4. Many students appeared to follow course recommendations from the objective test relating to enrollment in developmental writing or credit English.

5. Most students failed to follow course recommendations from the objective test relating to enrollment in developmental writing and credit English combinations.

6. Students who received a "2" on the objective test and did not enroll in a combination course tended to be evenly split between enrolling in developmental writing and credit English.

7. There are no statistically significant differences in writing sample scores due to the various topics or due to order of administering the objective test and writing sample.

8. For students enrolled in credit English, the objective test is the best predictor of course grade. Addition of other predictors adds no worthwhile increase to the correlation of predictors and course grade.

9. For students enrolled in developmental writing, the combination of age and writing sample score is the best predictor of course grade. Addition of other predictors adds no worthwhile increase to the correlation of predictors and course grade.

10. The possible use of the writing sample as an advisement tool should be based on its value of providing face validity rather than predictive validity.

11. The inter-rater reliability from two independent readings of selected writing samples was adequate but could be improved slightly with additional training of readers.

12. The objective test reliability was adequate but could be improved slightly through analysis of each item.

13. Readers were, in general, satisfied with the process for evaluating the writing samples.

14. Counselors were open to using the writing sample as part of the advisement process for entering students.

REFERENCES

Breland, H.M. & Gaynor, J.L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 1979, 16, 119-128.

Coffman, W.E. On the validity of essay tests of achievement. Journal of Educational Measurement, 1966, 3, 151-156.

Della-Piana, G., Odell, L., Cooper, C., & Endo, G. The writing skills decline: So what? Educational Technology, 1976, 16, 30-39.

Diederich, B. Measuring growth in English. Champaign, Illinois: National Council of Teachers of English, 1974.

Godshalk, F.I., Swineford, F., & Coffman, W.E. The measurement of writing ability. Princeton, N.J.: The College Board, 1966.

Olson, G.H., CORRAL. Unpublished ALGOL computer program. Dallas, Texas: Dallas Independent School District, 1979.

Olson, G.H., CROSSTAB. Unpublished ALGOL computer program. Dallas, Texas: Dallas Independent School District, 1978.

Roueche, J.E. & Archer, P.F. Entry level assessment in college. Community College Review, 1979, 6, 15-27.

Roueche, J.E. & Snow, J.J. Overcoming learning problems. San Francisco: Jossey-Bass Publishers, 1977.

ERIC Clearinghouse for Junior Colleges
8118 Math-Sciences Building
University of California
Los Angeles, California 90024

DEC 9 1983